不协调决策信息系统的知识约简及决策规则优化研究 *

汪 凌

(安庆师范大学 经济与管理学院, 安徽 安庆 246133)

摘　要：针对不协调决策信息系统的知识约简及决策规则的优化问题，引入分布约简和最大分布约简理论，提出一种基于分布区分对象集的知识约简算法，并得到具体的优化决策规则获取方法。该算法通过求解分布区分对象集和最小析取范式从而得到知识约简集，依据属性约简集挖掘出最优决策规则集。理论分析和实例结果表明该方法的有效性和实用性。

关键词：不协调决策信息系统；分布区分对象集；知识约简；决策规则

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2017.12.0849

Research on knowledge reduction and decision rules optimization in inconsistent decision information systems

Wang Ling

(*School of Economic & management*, *Anqing Normal University*, *Anqing Anhui* 246133, *China*)

Abstract: For the knowledge reduction and optimization decision rules of the inconsistent decision information system, by introducing distribution reduction and maximum distribution reduction theory, a knowledge reduction algorithm based on distribution set of objects is proposed in this paper, and approach optimal decision rule sets . The algorithm approach knowledge reduction sets by solving distribution set of objects and minimum disjunctive normal form, and mining optimal decision rules. Theoretical analysis and example results show that the method is effective and practical.

Key words: inconsistent decision information systems; distribution set of objects; knowledge reduction; decision rules

0　引言

粗糙集理论是一种处理不精确、不确定和噪声数据的有力工具，已被广泛应用于决策分析、信息处理、数据挖掘、模式识别等领域中[1, 2]。

不确定信息下的知识挖掘是粗集理论研究的重要问题之一。为了从不确定信息系统中挖掘更为优化的决策知识，关键需要对决策系统进行约简。目前，众多学者们对属性约简和知识挖掘理论开展了大量的研究工作。如张文修等在研究知识约简及决策规则挖掘时，引入了区分函数方法[3]，Kryszkiewicz[4]提出了相对约简概念，并给出具体的决策规则提取方法，文献[5]在分析集值信息系统的属性约简及知识获取问题时，基于容差关系和引入极大一致块方法等。上述研究大都面向协调信息系统进行的。由于数据增加、重要信息缺损、以及噪声干扰因素等，导致数据集不一致，大量决策信息系统都是不协调的。近年来，不协调决策信息系统得到广泛研究[6~10]。如 Kryszkiewicz 率先提出了分布约简和分配约简概念，但是尚未详细研究属性约简算法[11]。张文修等人[12]对前人的研究结论进行了拓展和延伸，引入差别矩阵，深入探讨了不协调决策系统属性约简算法。随着系统数据集的不断增加，这些算法复杂度逐步增大。为了解决这些算法中存在的问题，史德容等引入优势关系，研究了不协调区间值模糊序信息系统的分布约简和最大分布约简算法等[15]。如何高效从决策信息系统进行属性约简和获取最优化的规则知识，一直以来是研究人员共同面对的问题。

为此，本文在现有不协调决策信息系统属性约简方法基础上，通过引入分布约简和最大分布约简的概念，具体分析了不协调决策信息系统下的分布约简和最大分布约简理论，在此基础上，构造了一种基于分布区分对象集的属性约简算法，算法利用分布区分对象集的集对，计算最小析取范式得到属性约简结果，并挖掘出优化的决策规则集。最后，利用实例分析验证了算法的有效性和合理性。

1　不协调决策信息系统

定义 1[10,12]　设 (U,A,F,d) 为决策信息系统，$R_A=\{(x_i,x_j)\mid f_l(x_i)=f_l(x_j)(a_l\in A)\}$，$R_d=\{(x_i,x_j)\mid d(x_i)=d(x_j)\}$，若 $R_A\subseteq R_d$，则称 (U,A,F,d) 是协

收稿日期：2017-12-08；修回日期：2018-03-11　　基金项目：安徽省高校优秀人才重点支持计划资助项目（gxyqzd2016200）

作者简介：汪凌（1975-），男，副教授，博士（后），主要研究方向为决策科学与技术、不确定性理论及应用等（tcwling@126.com）.

调的，若 $R_A \not\subset R_d$，则称 (U,A,F,d) 是不协调的。

设 (U,A,F,d) 为决策信息系统，

$$R_B=\{(x_i,x_j)\mid f_l(x_i)=f_l(x_j)(a_l\in B)\}(B\subseteq A),$$

$$U/R_B=\{[x_i]_B\mid x_i\in U\},U/R_d=\{D_1,D_2,\cdots D_r\}$$

其中 $[x_i]_B=\{x_j\mid(x_i,x_j)\in R_B\}$。

对 $\forall x_i\in U$，记 $D(D_j/[x_i]_B)=\dfrac{|D_j\cap[x_i]_B|}{|[x_i]_B|}$ $(j\le r)$，则 U 上关于 B 的分布函数、最大分布函数、分配函数定义为

$$\mu_B(x_i)=(D(D_1/[x_i]_B),D(D_2/[x_i]_B),\cdots,D(D_r/[x_i]_B)),\quad(x_i\in U)\tag{1}$$

$$\eta_B(x_i)=\{D_{j0}\mid D(D_{j0}/[x_i]_B)\}=\max_{j\le r}D(D_j/[x_i]_B),\quad(x_i\in U)\tag{2}$$

$$\delta_B(x_i)=\{D_j\mid D_j\cap[x_i]_B\ne\varnothing\},(x_i\in U)\tag{3}$$

定义 2[12,14] 设 (U,A,F,d) 为决策信息系统，$B\subseteq A$。

a）对 $\forall x_i\in U$，$\exists\ \mu_B(x_i)=\mu_A(x_i)$，则 B 是分布协调集，如果 B 的任何真子集都不是分布协调集，则必定是分布约简集。

b）对 $\forall x_i\in U$，$\exists\ \eta_B(x_i)=\eta_A(x_i)$，则 B 是最大分布协调集，如果 B 的任何真子集都不是最大分布协调集，则必定是最大分布约简集。

c）对 $\forall x_i\in U$，$\exists\ \delta_B(x_i)=\delta_A(x_i)$，则 B 是分配协调集，如果 B 的任何真子集都不是分配协调集，则必定是分配约简集。

由定义 1 和 2 可知，分布协调集指决策类上分布函数保持不变的属性集，最大分布协调集指最大分布决策类保持不变的属性集，而分配协调集指所有对象的决策类保持不变。

定理 1 设 (U,A,F,d) 为不协调决策信息系统，$B\subseteq A$，如果 B 是分布协调集，则必定是最大分布协调集。

证明 假设 B 是 (U,A,F,d) 的分布协调集，根据定义 2 的 a），对 $\forall x_i\in U$，$B\subseteq A$，$\exists\ \mu_B(x_i)=\mu_A(x_i)$ $(x_i\in U)$，有 $D(D_j/[x_i]_B)=D(D_j/[x_i]_A)$ $(\forall j\le r)$ 成立。由定义 2 的 b）可知，$\exists\ \eta_B(x_i)=\eta_A(x_i)$ $(x_i\in U)$ 成立，因此 B 是最大分布协调集。

定理 2 设 (U,A,F,d) 为不协调决策信息系统，$B\subseteq A$，如果 B 是分布协调集，则 B 必定是分配协调集。

证明 一方面，假设 B 是分布协调集，则对 $\forall x_i\in U$，$B\subseteq A$，根据分不协调集的定义，有 $\mu_B(x_i)=\mu_A(x_i)$，即 $D(D_j/[x_i]_B)=D(D_j/[x_i]_A)$，$(\forall j\le r)$。

另一方面，假设 $D_j\in\delta_B(x_i)$，则由分配函数的定义，可知 $D_j\cap[x_i]_B\ne\varnothing$，因此 $D(D_j/[x_i]_B)\ne0$，于是 $D(D_j/[x_i]_A)\ne0$，从而 $D_j\cap[x_i]_A\ne\varnothing$。因此，$D_j\in\delta_A(x_i)$，于是就有 $\delta_A(x_i)\supseteq\delta_B(x_i)$。而当 $B\subseteq A$ 时，$R_B\supseteq R_A$，对 $\forall x_i\in U$，存在 $[x_i]_B\supseteq[x_i]_A$，即 $\delta_B(x_i)=\delta_A(x_i)$，$\forall x_i\in U$，因此 B 一定是分布协调集。

例 1 决策信息系统如表 1 所示。

按照分类等价关系有：

$$U/R_A=\{\{x_1,x_2\},\{x_3,x_4,x_5\},\{x_6\}\},$$

$$U/R_d=\{\{x_1,x_2,x_3,x_4\},\{x_5,x_6\}\}$$

表 1 决策信息系统

U	a_1	a_2	d
x_1	1	1	1
x_2	1	1	1
x_3	1	2	1
x_4	1	2	1
x_5	1	2	2
x_6	2	2	2

显然，$R_A\not\subset R_d$，因此 (U,A,F,d) 为不协调决策信息系统，并且

$$R_{a_1}=\{\{x_1,x_2,x_3,x_4,x_5\},\{x_6\}\},$$

$$R_{a_2}=\{\{x_1,x_2\},\{x_3,x_4,x_5,x_6\}\},$$

于是得到决策分布函数为：

$\mu_A(x_1)=\mu_A(x_2)$ =(1,0), $\mu_A(x_3)=\mu_A(x_4)=\mu_A(x_5)$ =(0.67,0.33), $\mu_A(x_6)$ =(0,1);

$\mu_{a_1}(x_1)=\mu_{a_1}(x_2)=\mu_{a_1}(x_3)=\mu_{a_1}(x_4)=\mu_{a_1}(x_5)$ =(0.8,0.2), $\mu_{a_1}(x_6)$ =(0,1);

$\mu_{a_2}(x_1)=\mu_{a_2}(x_2)$ =(1,0), $\mu_{a_2}(x_3)=\mu_{a_2}(x_4)=\mu_{a_2}(x_5)=\mu_{a_2}(x_6)$ =(0.5,0.5);

最大分布函数为 $\eta_A(x_1)=\eta_A(x_2)=\eta_A(x_3)=\eta_A(x_4)=\eta_A(x_5)=\{D_1\}$，$\eta_A(x_6)=\{D_2\}$

$\eta_{a_1}(x_1)=\eta_{a_1}(x_2)=\eta_{a_1}(x_3)=\eta_{a_1}(x_4)=\eta_{a_1}(x_5)=\{D_1\},\eta_{a_1}(x_6)=\{D_2\}$

$\eta_{a_2}(x_1)=\eta_{a_2}(x_2)=\{D_1\}$, $\eta_{a_2}(x_3)=\eta_{a_2}(x_4)=\eta_{a_2}(x_5)=\eta_{a_2}(x_6)=\{D_1,D_2\}$

由定义 2 可知，A 属于分布约简集，但不是最大分布约简集。

2 知识约简理论分析

不协调决策信息系统 (U,A,F,d)，对 $\forall x_i\in U$，根据上述有关定义和定理，可以准确区分属性子集的约简性和协调性。下面首先给出分布区分对象集对概念及其性质定理。

定义 3[12,14] 设 (U,A,F,d) 为不协调决策信息系统，$U/R_A=\{[x_1]_A,[x_2]_A,\cdots,[x_i]_A\}$，$\mu_A(x_i)=(D(D_1/[x_i]_A),D(D_2/[x_i]_A),\cdots,D(D_r/[x_i]_A))$ $(\forall x_i\in U)$，$\eta_A(x_i)=\{D_{j0}\mid D(D_{j0}/[x_i]_A)=\max_{j\le r}D(D_j/[x_i]_A)\}$，其分布区分对象集对、最大分布区分对象集对分别定义为

$$D_\mu^*=\{([x_i]_A,[x_j]_A)\mid\mu_A(x_i)\ne\mu_A(x_j)\}\tag{4}$$

$$D_\eta^*=\{([x_i]_A,[x_j]_A)\mid\eta_A(x_i)\ne\eta_A(x_j)\}\tag{5}$$

定义 4[12,15] 设 (U,A,F,d) 为不协调决策信息系统，$U/R_A=\{C_1,C_2,\cdots,C_m\}$，$f_{a_k}(C_i)$ 表示 a_k 关于 C_i 的值。

a）记

$$D_1(C_i,C_j)=\begin{cases}\{a_k\mid a_k\in A,f_{a_k}(C_i)\ne f_{a_k}(C_j)\} & (C_i,C_j)\in D_1^*\\ \varnothing & (C_i,C_j)\notin D_1^*\end{cases}\tag{6}$$

则定义 $M_\mu=\{D_\mu(C_i,C_j),i,j\le m\}$ 为分布区分矩阵。

b）记

$$D_\eta(C_i,C_j)=\begin{cases}\{a_k \mid a_k\in A, f_{a_k}(C_i)\neq f_{a_k}(C_j) & (C_i,C_j)\in D_\eta^*\\ \varnothing & (C_i,C_j)\notin D_\eta^*\end{cases}\quad(7)$$

则定义 $M_\eta=\{D_\eta(C_i,C_j),i,j\le m\}$ 为最大分布区分矩阵。

显然，定义 4 是在定义 2 基础上的进一步拓展和延伸。由上述定义易知下列性质和定理。

性质 1　设 (U,A,F,d) 为不协调决策信息系统，$U/R_A=\{C_1,C_2,\cdots,C_m\}$，易知 M_μ、M_η 具有以下性质：

a）M_μ、M_η 为对称矩阵，即 $D_\mu(C_i,C_j)=D_\mu(C_j,C_i)$，$D_\eta(C_i,C_j)=D_\eta(C_j,C_i)$，$\forall i,j\le m$；

b）M_μ、M_η 对角线上元素都为 A，即 $D_\mu(C_i,C_i)=D_\eta(C_i,C_i)=A$，$\forall i\le m$；

c）$D_\mu(C_i,C_j)\subseteq D_\mu(C_i,C_k)\cup D_\mu(C_k,C_j)$，$\forall i,k,j\le m$

$D_\eta(C_i,C_j)\subseteq D_\eta(C_i,C_k)\cup D_\eta(C_k,C_j)$，$\forall i,k,j\le m$

定理 3　设 (U,A,F,d) 为不协调决策信息系统，对于 $\forall B\subseteq A$，则有：

a）B 为分布协调集当且仅当对 $\forall (C_i,C_j)\in D_\mu^*$，$\exists B\cap D_\mu(C_i,C_j)\neq\varnothing$。

b）B 为最大分布协调集当且仅当对 $\forall (C_i,C_j)\in D_\eta^*$，$\exists B\cap D_\eta(C_i,C_j)\neq\varnothing$。

证明　a）充分性。假设 B 为分布协调集，若 $\forall (C_i,C_j)\in D_\mu^*$，记 $C_i=[x_i]_A$，$C_j=[x_j]_A$，则由分布协调集的定义，知 $\mu_A(x_i)\neq\mu_A(x_j)$。根据定义 2 的 a）可知 $[x_i]_A\cap[x_j]_A=\varnothing$。则 $\exists a_k\in B$，使得 $f_k(x_i)\neq f_k(x_j)$，即 $f_k(C_i)\neq f_k(C_j)$，因而有 $a_k\in D_\mu(C_i,C_j)$ 成立。从而，结论 $B\cap D_\mu(C_i,C_j)\neq\varnothing$ 成立。

必要性。假设 $\exists (C_i,C_j)\in D_\mu^*$，使得 $B\cap D_\mu(C_i,C_j)=\varnothing$ 成立，记 $C_i=[x_i]_A$，$C_j=[x_j]_A$，则由分布协调集的定义，可知 $\mu_A(x_i)\neq\mu_A(x_j)$。对于 $\forall a_k\in B$，必然 $\exists a_k\notin D_\mu(C_i,C_j)$，使得 $f_k(C_i)=f_k(C_j)$，即 $f_k(x_i)=f_k(x_j)$，这说明 $[x_i]_B=[x_j]_B$，根据定义 2 的 a）可知 B 不为分布协调集。证毕。

b）充分性。假设 B 为最大分布协调集，若 $\forall (C_i,C_j)\in D_\eta^*$，记 $C_i=[x_i]_A$，$C_j=[x_j]_A$，则由分不协调集的定义可知 $\mu_A(x_i)\neq\mu_A(x_j)$。根据定义 2 的 b）可知 $[x_i]_A\cap[x_j]_A=\varnothing$。则 $\exists a_k\in B$，使得 $f_k(x_i)\neq f_k(x_j)$，即 $f_k(C_i)\neq f_k(C_j)$，因而有 $a_k\in D_\eta(C_i,C_j)$ 成立。从而，结论 $B\cap D_\eta(C_i,C_j)\neq\varnothing$ 成立。

必要性。假设 $\exists (C_i,C_j)\in D_\eta^*$，使得 $B\cap D_\eta(C_i,C_j)=\varnothing$ 成立，记 $C_i=[x_i]_A$，$C_j=[x_j]_A$，则根据分布协调集的定义，可知 $\mu_A(x_i)\neq\mu_A(x_j)$。对于 $\forall a_k\in B$，必然 $\exists a_k\notin D_\eta(C_i,C_j)$，使得 $f_k(C_i)=f_k(C_j)$，即 $f_k(x_i)=f_k(x_j)$，这说明 $[x_i]_B=[x_j]_B$，根据定义 2 的 b）可知，B 不为最大分布协调集。证毕。

定理 3 实质上给出了判定不协调决策信息系统属性子集是否协调的理论依据。

定义 5[12,15]　设 (U,A,F,d) 为不协调决策信息系统，$U/R_A=\{C_1,C_2,\cdots,C_m\}$，则其分布区分公式、最大分布区分公式分别定义为

$$F_\mu=\bigwedge_{(C_i,C_j)\in D_\mu^*}\{\vee D_\mu(C_i,C_j)\}\quad(8)$$

$$F_\eta=\bigwedge_{(C_i,C_j)\in D_\eta^*}\{\vee D_\eta(C_i,C_j)\}\quad(9)$$

定理 4　设 (U,A,F,d) 为不协调决策信息系统，F_μ 最小析取范式简记为 $F_\mu=\bigvee_{k=1}^{m}\{\bigwedge_{i=1}^{n}a_i\}$，若 $B_k=\{a_i\mid i=1,2,\cdots,n\}$，则 $\{B_k\mid k=1,2,\cdots,m\}$ 包含所有分布约简集。

证明　对 $\forall k\le m$，$(C_i,C_j)\in D_\mu^*$，由最小析取范式的定义知 $B_k\cap D_\mu(C_i,C_j)\neq\varnothing$，根据定理 3a）可知 B_k 为分布协调集。若 $F_\mu=\bigvee_{k=1}^{m}(B_k)$ 从 B_k 中任意约去元素 a_i 成为 $B_k^{'}$，即 $B_k^{'}=B_k-\{a_i\}$，则 $\exists x_i,x_j\in U$，且 $C_i=[x_i]_A$，$C_j=[x_j]_A$，$(C_i,C_j)\in D_\mu^*$，使 $B_k^{'}\cap D_\mu(C_i,C_j)=\varnothing$，则 $B_k^{'}$ 不是分布协调集，从而 B_k 是分布约简。而区分公式 F_μ 包括所有 $D_\mu(C_i,C_j)$，因此不含其他分布约简。

定理 5　设 (U,A,F,d) 为不协调决策信息系统，F_η 最小析取范式简记为 $F_\eta=\bigvee_{k=1}^{m}\{\bigwedge_{i=1}^{n}a_i\}$，若 $B_k=\{a_i\mid i=1,2,\cdots,n\}$，则 $\{B_k\mid k=1,2,\cdots,m\}$ 包含所有最大分布约简集。

证明　对 $\forall k\le m$，$(C_i,C_j)\in D_\eta^*$，由最小析取范式的定义知 $B_k\cap D_\eta(C_i,C_j)\neq\varnothing$，根据定理 3（2），可知 B_k 为最大分布协调集。若 $F_\eta=\bigvee_{k=1}^{m}(B_k)$ 从 B_k 中任意约去元素 a_i 成为 $B_k^{'}$，即 $B_k^{'}=B_k-\{a_i\}$，则 $\exists x_i,x_j\in U$，且 $C_i=[x_i]_A$，$C_j=[x_j]_A$，$(C_i,C_j)\in D_\eta^*$，使 $B_k^{'}\cap D_\eta(C_i,C_j)=\varnothing$，则 $B_k^{'}$ 不是最大分布协调集，从而 B_k 是最大分布约简。而最大区分公式 F_η 包括所有 $D_\eta(C_i,C_j)$，因此不含其他最大分布约简。

定理 4 和 5 分别给出了利用分布区分公式和最大分布区分公式，计算其最小析取范式，可以确定不协调决策信息系统各种相对约简集的方法。

3　不协调决策信息系统的知识约简方法

3.1　算法原理与描述

根据知识约简相关理论分析，一方面，由定义 3 和 4 可知，可以计算不协调决策信息系统的分布区分对象集对和最大分布区分对象集对，从而准确区分对象属性集;另一方面，根据定义 5、定理 4 和 5，可以构建分布区分矩阵和最大分布区分矩阵，通过计算最小析取范式，得出所有相对约简集。依据上述有关定义及性质定理，设计一种基于分布区分对象集的知识约简算法，具体算法过程描述如下：

算法 1　基于分布区分对象集的知识约简算法

输入：不协调决策信息系统 $S=(U,A,F,d)$

输出：属性约简集 Re*duction*

a)对于 $S=(U,A,F,d)$ 中 $\forall x_i\in U$，计算 $[x_i]_A$；

b)对于 $S=(U,A,F,d)$ 中 $\forall x_i\in U$，计算决策等价类 $[x_i]_d$；

$U/R_d=\{D_1,D_2,\cdots,D_r\}=\{(x_i,x_j)\in U\times U\mid d(x_i)=d(x_j)\}$；

c)计算 $S=(U,A,F,d)$ 的分布约简集；

(a)根据定义 1，对 $\forall x_i\in U$，求解 $\mu_A(x_i)$：

$\mu_A(x_i)=(D(D_1/[x_i]_A),D(D_2/[x_i]_A),\cdots,D(D_r/[x_i]_A))$；

(b) 计算 D_μ^* 得分布区分对象集的集对 D_μ^*：$D_\mu^*=\{([x_i]_A,[x_j]_A)\mid \mu_A(x_i)\neq\mu_A(x_j)\}$；

(c)根据分布区分对象集的集对计算分布区分属性集 $D_\mu(C_i,C_j)$：

$$D_\mu(C_i,C_j)=\begin{cases}\{a_k\mid a_k\in A,f_{a_k}(C_i)\neq f_{a_k}(C_j) & (C_i,C_j)\in D_\mu^*\\ \varnothing & (C_i,C_j)\notin D_\mu^*\end{cases};$$

(d) 根据 $D_\mu(C_i,C_j)$ 计算分布区分矩阵 M_μ：

$$M_\mu=\wedge_{i,j}\{\vee D_\mu(C_i,C_j)\}=\wedge_{(C_i,C_j)\in D_\mu^*}\{\vee D_\mu(C_i,C_j)\};$$

(e)计算 $F_\mu=\vee_{k=1}^{m}\{\wedge_{i=1}^{n}a_i\}$ 的最小析取范式，得分布约简集 $F_\mu^{\min}$；

d)计算 $S=(U,A,F,d)$ 的最大分布约简集；

(a)根据定义 1，对于 $\forall x_i\in U$，求解 $\eta_A(x_i)$：

$$\eta_A(x_i)=\{D_{j0}\mid D(D_{j0}/[x_i]_A)=\max_{j\le r}D(D_j/[x_i]_A)\},$$

$$(x_i\in U);$$

(b) 计算 D_η^* 得最大分布区分对象集的集对 D_η^*：$D_\eta^*=\{([x_i]_A,[x_j]_A)\mid \eta_A(x_i)\neq\eta_A(x_j)\}$；

(c)根据最大分布区分对象集的集对计算最大分布区分属性集 $D_\eta(C_i,C_j)$：

$$D_\eta(C_i,C_j)=\begin{cases}\{a_k\mid a_k\in A,f_{a_k}(C_i)\neq f_{a_k}(C_j) & (C_i,C_j)\in D_\eta^*\\ \varnothing & (C_i,C_j)\notin D_\eta^*\end{cases};$$

(d)根据 $D_\eta(C_i,C_j)$ 计算最大分布区分矩阵

$$M_\eta=\wedge_{i,j}\{\vee D_\eta(C_i,C_j)\}=\wedge_{(C_i,C_j)\in D_\eta^*}\{\vee D_\eta(C_i,C_j)\};$$

(e)计算 $F_\eta=\vee_{k=1}^{m}\{\wedge_{i=1}^{n}a_i\}$ 的最小析取范式，得所有最大分布约简集 $F_\eta^{\min}$；

e)输出 $F_\mu^{\min}$ 和 $F_\eta^{\min}$。

3.2　算法分析

1）时间复杂度分析

步骤 a)和 b)的时间开销为 $O(|U|)$。对于步骤 c)中,(a)(b)可在时间复杂度 $O(|U|^2)$ 内完成;（c）时间复杂度为 $O(|D_\mu^*|.|C|)$；(d)(e)时间复杂度接近于 $O(|U|^2)$；步骤 d)中，(a)的时间复杂度为 $O(|\mu_A(x)|)$；(b)的时间复杂度为 $O(|U|^2)$；（c）的时间复杂度为 $O(|D_\eta^*|.|C|)$；(e)(f)可在时间复杂度 $O(|U|^2)$ 内完成。因此，本算法总的时间复杂度为 $\max\{O(|U|^2),O(|D_\mu^*|.|C|)\}$。

2）空间复杂度分析

算法 1 的空间复杂度为 $\max\{O(|U||C|),O(|D_\mu^*|.|C|)\}$。

与算法 1 相比，文献[10]中采用的属性约简算法的时间复杂度取决于广义分配辨识矩阵约简算法的时间复杂度，为 $O(|C|^3.|U|^2)$，计算过程较为复杂。文献[14]在计算不协调信息系统的属性约简时，需要构建决策系统的区分矩阵来存储差别属性，因而占用较大的存储空间，即算法的空间复杂度为 $O(|C|.|U|^2)$。显然，与目前常用的属性约简算法相比较，本文算法在时间复杂度和空间复杂度上都具有较大的优势，并且能够求出系统所有分布约简集和最大分布约简集。

4　基于分布约简算法的决策规则获取

不协调决策信息系统的决策规则获取是在属性约简基础上归纳出条件属性和决策属性之间的关联关系。在决策规则获取过程中，决策规则的前件表示属性集描述，后件对应决策结论。设 $S=(U,A,F,d)$ 为不协调决策信息系统，$B\subseteq A$，$d=\{d\}$，$U/R_d=\{d_1,d_2,\cdots d_r\}$，则决策规则可表示为：

$r_x:\wedge(a,v)\to\vee(d,w)$,其中，$a\in B\subseteq A$，$v\in V_a$，$w\in V_d$，$V_a$ 表示 a 的值域，V_d 表示 d 的有限值域。

r_x 的可信度因子用 $C(r_x^i)$ 衡量

$$C(r_x^i)=\max_{i=1}^{m}\{\frac{|[x]_B\cap[x]_d|}{|[x]_B|}\}$$

其中:$[x]_B$ 为 B 的等价类，$[x]_d$ 为 d 的等价类，|?| 为集合基数。易知，$C(r_x)=1$，表明 r_x 是确定规则；$0<C(r_x)<1$，表明 r_x 是不确定规则。

定义 6[16~18]　设 $\wedge(a,v)\to\vee(d,w)$ 为一般决策规则，则称 $\mathrm{Re}\,d(\wedge(a,v))\to\vee(d,w)$ 为优化决策规则。此定义表明，优化决策规则是确定性程度最高且条件属性描述最简洁的决策规则。

4.1　不协调决策信息系统优化决策规则获取算法

根据不协调决策信息系统的知识约简和决策规则获取理论分析，从不协调决策系统中提取决策规则就是利用数据的分布约简和最大分布约简构造新的决策表，再引入规则可信度，归纳出优化的决策规则。从不协调决策信息系统中获取优化决策规则具体算法如下：

算法 2　基于分布约简集的优化决策规则获取算法

输入：不协调决策信息系统 $S=(U,A,F,d)$。

输出：最优决策规则集。

a)根据算法 1 给出的约简算法，计算 $S=(U,A,F,d)$ 中所有分布约简集 $F_\mu^{\min}$ 和最大分布约简集 $F_\eta^{\min}$；

b) 分别根据 $F_\mu^{\min}$ 和 $F_\eta^{\min}$，构建新决策信息系统 $S'=(U,A,F,d)$；

c)规则获取与选择。计算可信度，当可信度达到设定的阈值时，提取 $S'=(U,A,F,d)$ 中所有符合条件的决策规则；

d)获取最优决策规则。将最后得到的决策规则进行整合，分别输出 $C(r_x)=1$ 的确定规则集和 $0<C(r_x)<1$ 的不确定规则集。

5　算例分析

为了进一步分析本文算法的有效性，表 2 给出了一个不协调决策信息系统 $S=(U,A,F,d)$，其中 $U=\{x_1,x_2,\cdots,x_6\}$，$C=\{a_1,a_2,a_3,a_4\}$，$d=\{d\}$。

表 2　不协调决策信息系统

U	a_1	a_2	a_3	a_4	d
x_1	1	2	2	2	1
x_2	2	1	1	2	2
x_3	2	1	2	2	1
x_4	2	1	2	2	1
x_5	2	1	1	1	2
x_6	2	1	2	2	2

a）根据算法 1 的步骤 a)，可计算条件属性等价类为 $C_1=[x_1]_A=\{x_1\}$；$C_2=[x_2]_A=\{x_2\}$；$C_3=[x_3]_A=\{x_3,x_4,x_6\}$；$C_4=[x_5]_A=\{x_5\}$

b）根据算法 1 的步骤 b)，计算决策属性等价类为：

$$d_1=\{x_1,x_3,x_4\}\,;\,d_2=\{x_2,x_5,x_6\}$$

c）根据算法 1 的步骤 c)计算分布约简：

(a)根据步骤 c)中（a）计算分布函数 $\mu_A(x_i)$ 的值，$i=1,2,\cdots,6$，$\mu_A(x_1)=(1,0)$；$\mu_A(x_2)=(0,1)$；$\mu_A(x_3)=\mu_A(x_4)=\mu_A(x_6)=(0.67,0.33)$；$\mu_A(x_5)=(0,1)$；

(b)根据步骤 c)中（b）计算分布区分对象集对：$D_\mu^*=\{(C_1,C_2),(C_1,C_3),(C_1,C_4),(C_2,C_3),(C_3,C_4)\}$

(c)根据步骤 c)中（c）计算分布区分属性集：

$$D_\mu(C_1,C_2)=\{a_1,a_2,a_3\}\,;\,D_\mu(C_1,C_3)=\{a_1,a_2\}\,;$$
$$D_\mu(C_1,C_4)=\{a_1,a_2,a_3,a_4\}\,;\,D_\mu(C_2,C_3)=\{a_3\}\,;$$
$$D_\mu(C_3,C_4)=\{a_3,a_4\}$$

(d)根据步骤 c)中(e)(f)，计算最小析取范式：

$$\begin{aligned}F_\mu=(a_1\vee a_2\vee a_3)\wedge(a_1\vee a_2)\wedge\\(a_1\vee a_2\vee a_3\vee a_4)\wedge a_3\wedge(a_3\vee a_4)\end{aligned}=(a_1\wedge a_3)\vee(a_2\wedge a_3)$$

因此，$S=(U,A,F,d)$ 的分布约简集为 $\{a_1,a_3\}$ 和 $\{a_2,a_3\}$。

d）根据算法 1 的步骤 d)求最大分布约简：

(a)根据步骤 d)中(a)计算最大分布函数 $\eta_A(x_i)$ 的值，$i=1,2,\cdots,6$

$$\eta_A(x_1)=\{D_1\}\,;\,\eta_A(x_2)=\{D_2\}\,;$$
$$\eta_A(x_3)=\eta_A(x_4)=\eta_A(x_6)=\{D_1\}\,;\,\eta_A(x_5)=\{D_2\}$$

(b)根据步骤 d)中(b)计算最大分布区分对象的集对：

$$D_\eta^*=\{(C_1,C_2),(C_1,C_4),(C_2,C_3),(C_3,C_4)\}$$

(c)根据步骤步骤 d)中(c)计算最大分布区分属性集：

$$D_\eta(C_1,C_2)=\{a_1,a_2,a_3\}\,;\,D_\eta(C_1,C_4)=\{a_1,a_2,a_3,a_4\}\,;$$
$$D_\eta(C_2,C_3)=\{a_3\}\,;\,D_\eta(C_3,C_4)=\{a_3,a_4\}$$

(d)根据步骤 d)中(d)和(e)，计算最小析取范式：

$$\begin{aligned}F_\eta=(a_1\vee a_2\vee a_3)\wedge\\(a_1\vee a_2\vee a_3\vee a_4)\wedge a_3\wedge(a_3\vee a_4)\end{aligned}=a_3$$

因此，$S=(U,A,F,d)$ 的最大分布约简为 $\{a_3\}$。

e）根据算法 2，基于 $F_\mu^{\min}=\{\{a_1,a_3\}\}$，对应优化决策规则集为

$$r_1:(a_1,1)\wedge(a_3,2)\rightarrow(d,1)$$
$$r_2:(a_1,2)\wedge(a_3,1)\rightarrow(d,2)$$
$$r_3:(a_1,2)\wedge(a_3,2)\rightarrow(d,1)\vee(d,2)$$

基于 $F_\mu^{\min}=\{\{a_2,a_3\}\}$，对应优化决策规则集为

$$r_1:(a_2,2)\wedge(a_3,2)\rightarrow(d,1)$$
$$r_2:(a_2,1)\wedge(a_3,1)\rightarrow(d,2)$$
$$r_3:(a_2,1)\wedge(a_3,2)\rightarrow(d,1)\vee(d,2)$$

f）基于 $F_\eta^{\min}=\{\{a_3\}\}$，对应优化决策规则集为

$$r_1:(a_3,1)\rightarrow(d,2)$$
$$r_2:(a_3,2)\rightarrow(d,1)\vee(d,2)$$

因此，不协调决策信息系统 $S=(U,A,F,d)$ 的分布约简集 $B_1=\{\{a_1,a_3\},\{a_2,a_3\}\}$，最大分布约简集 $B_2=\{a_3\}$，并且都能够从决策系统中挖掘出更为优化的决策规则集，从而使决策知识更加科学合理。

6　结束语

本文面向不协调决策信息系统，引入分布约简和最大分布约简概念，获取系统相对约简集，构建了基于分布区分对象集的知识约简算法，该算法通过求解分布区分对象集集对和最小析取范式从而得到知识约简集，并能充分挖掘优化决策规则集，提高了决策规则获取算法效率，为不协调信息系统决策知识发现提供了一种新的思路与方法，实例验证了本文方法的有效性和实用性。下一步将针对系统数据变化情况，重点研究知识动态挖掘及其应用问题。

参考文献：

[1] Pawlak Z, Skowron A. Rudiments of rough sets [J]. Information Sciences, 2007, 177 (1): 3-27.

[2] Estaji A A, Hooshmandasl M R, Davva B. Rough set theory applied to lattice theory [J]. Information Science, 2012, 200: 108-122.

[3] Zhang W X, Liang Y, Wu W Z. Information systems and knowledge discovery [M]. Beijing: Science Press, 2003: 22-47.

[4] Kryszkiewicz M. Rules in incomplete information systems [J]. Information Science, 1999, 113: 271-292.

[5] Guang Yanyong, Wang Hongkai. Set-valued information systems [J]. Information Science, 2006, 176 (5): 2507-2525.

[6] Chen D G, Wang C Z, Hu Q H. A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets [J]. Information Sciences, 2007, 177 (17): 3500-3518.

[7] Leung Y, Ma J M, Zhang W X, Li T J. Dependence-space-based attribute reductions in inconsistent decision information systems [J]. International Journal of Approximate Reasoning, 2008, 49 (3): 623-630.

[8] Miao D Q, Zhao Y, Yao Y Y, *et al.* Relative reducts in consistent and inconsistent decision tables of the pawlak rough set model [J]. Information Sciences, 2009, 179 (24): 4140-4150.

[9] 张文修，仇国芳. 基于粗糙集的不确定决策 [M]. 北京：清华大学出版社, 2005.

[10] 莫京兰，翁世洲，吕跃进. 不协调序信息系统的广义分配约简 [J]. 模

糊系统与数学, 2014, 28 (6): 163-168.

[11] Kryszkiewicz M. Comparative studies of alternative type of knowledge reduction in inconsistent systems [J]. International Journal of Intelligent Systems, 2001, 16 (1): 105-120.

[12] 张文修, 米据生, 吴伟志. 不协调目标信息系统的知识约简 [J]. 计算机学报, 2003, 26 (1): 12-18.

[13] 桑彬彬, 徐伟华. 直觉模糊序决策信息系统的分配约简 [J]. 计算机科学, 2017, 44 (6): 75-79.

[14] 方莲花, 李克典. 基于优势—等价关系下不协调目标信息系统的分布约简 [J]. 模糊系统与数学, 2013, 27 (3): 182-189.

[15] 史德容, 徐伟华. 区间值模糊决策序信息系统的分布约简 [J]. 计算机科学与探索, 2017, 11 (4): 652-658.

[16] 郭庆, 吴磊. 多粒度背景下直觉模糊信息系统的粗糙集及其决策 [J]. 系统工程与电子技术, 2016, 38 (2): 347-351.

[17] 吴磊, 杨善林, 郭庆. 优势关系下直觉模糊目标信息系统的上近似约简 [J]. 模式识别与人工智能, 2014, 27 (4): 300-304.

[18] 郭庆, 戴习民. 区间值信息系统的多粒度粗糙集及其决策 [J]. 合肥工业大学学报, 2017, 40 (2): 284-287.

[19] 刘芳, 李天瑞. 基于边界域的不完备信息系统属性约简算法 [J]. 计算机科学, 2016, 43 (3): 242-245.

[20] 鲍中奎, 杨善林. 直觉模糊目标信息系统的知识约简 [J]. 中国科学技术大学学报, 2015, 45 (9): 776-783.